Endogenous Games and Mechanisms: Side Payments Among Players

Matthew O. Jackson and Simon Wilkie*

September 2000
This Draft: July 3, 2002

Abstract

We characterize the outcomes of games when players may make binding offers of strategy contingent side payments before the game is played. This does not always lead to efficient outcomes, despite complete information and costless contracting. The characterizations are illustrated in a series of examples, including voluntary contribution public good games, Cournot and Bertrand oligopoly, principal-agent problems, and commons games, among others.

*Division of Humanities and Social Sciences, 228- 77, California Institute of Technology, Pasadena, California 91125, USA, jacksonm@hss.caltech.edu and wilkie@bondi.caltech.edu. We thank Ken Hendricks, Philippe Jéhiel, Ehud Kalai, Roger Lagunoff, Bentley MacLeod, Nolan Miller, Hakan Orbay, Mike Peters, and seminar participants at the University of Arizona, Caltech, University of Texas, University of Toronto, U.B.C., USC, and the Decentralization Conference for helpful comments and suggestions. Financial support under NSF grants SES-9986190 and SES-9986676 is gratefully acknowledged.

1 Introduction

Game theory and mechanism design are powerful tools that have become essential in the modeling of economic interactions. Generally, in modeling interactions from public goods contributions to imperfect competition among firms, the game being played or mechanism being designed is viewed in isolation. That is, we usually treat the game as being fixed from the players' perspective. The analysis of many games viewed in such isolation leads to a prediction of an inefficient outcome, since in many contexts there are externalities present. For instance voluntary public goods contributions games and commons games have well-known free rider problems and equilibria that are Pareto inefficient. Similar results hold for many other games, such as those with imperfect competition or production externalities such as pollution.

In practice, however, we often see players side contracting to improve efficiency. For instance, large donors often match donations of other donors in contributions games. We see this in public radio and television station fundraising where one donor will agree to donate an amount equal to that donated by in some time period (sometimes even subject to minimum or maximum donations or subject to the donations clearing some hurdle). In fact this is now generally a standard practice. For example, many employers now offer to match their employees' contributions to any charities. On an intuitive level this type of side contracting can help overcome externalities and reduce inefficiencies. The promise to match donations increases the impact that a donation has and can essentially compensate for the externality — representing the value that the donation would have to others. Similar side contracting appears in the tragedy of the commons games in the form international fishing and international pollution agreements, where often some promises of side payments are included. Again, the side payments can help promote efficiency by changing the incentives so that each party more fully sees the total impact or value that its actions generate.

While one can see an intuitive role for such side contracting, it is important to fully understand how such side contracting affects the outcome of the game. Which side contracts will agents write, and will the ability of agents to side contract lead to efficiency? These are the central questions that we address in this paper. There is a widespread belief among economists in the efficiency properties of what may be called "Coasian Contracting." The simple but powerful idea put forth by Coase (1960) says that if property rights are well-defined, and bargaining is costless, then rational agents faced with externalities should contract to come to an efficient outcome. Roughly speaking, with fully symmetric information and no transactions costs, agents should be able to come to an agreement that supports an efficient strategy profile as a equilibrium point of the game with side payments. In this paper we hold this reasoning to a careful scrutiny, and find that the issue is surprisingly subtle. Side contracting does not always lead to efficiency even when there are no transactions costs, complete information, and binding contracts. In fact, even if we start with a game that has Pareto efficient Nash equilibria, side contracting on the part of players can change the equilibrium structure so that all equilibria are inefficient!

The perspective we take here is to view a game as being embedded in a larger game where in a first stage players may engage in side contracting that can effectively rewrite payoff functions and then play the eventual altered game in the second period. This takes the eventual game that is played to be *endogenous*. In particular, we examine the following scenario: A set of agents are to play a game. The payoffs of the game are known to all players (i.e., common knowledge). Before playing the game, the agents can make enforceable offers of strategy contingent side payments to each other. So, players can make offers of the sort, "If actions x are played in the game we are about to play, I will pay you an amount y." The offers that can be made can be contingent on the actions of more than one player and can offer different payments for different combinations of actions. The offers are publicly observed and legally enforceable, and actions taken in any subsequent play of the game are also observable to any third party such as a court. If any such offers are made, then the net payoffs in the game to be played are modified and this affects the equilibrium behavior. From this point of view, the game has become endogenous. We explore how the ability to make such enforceable strategy contingent offers affects the equilibrium payoffs of the game.

Our main results are a complete characterization of the set of supportable equilibrium payoffs in endogenous games. We show that the equilibrium outcomes of a game with this costless stage of pre-play contracting need not be efficient. We cannot completely rely on endogenous side contracting to solve the inefficiency problem (and moreover, side contracting may introduce inefficiency where it was not a problem before). Our results provide a complete characterization of the supportable equilibrium outcomes, and how these depend on the structure of the game. Thus, we identify the class of games for which such endogenous side contracting will be efficient. This class includes some interesting examples such as some specifications of the tragedy of the commons and Bertrand games, but also excludes many interesting examples like voluntary contribution public good games and Cournot games.

The intuition behind the results, and the reason that efficiency is not always obtained, is that players wish to use side contracting to modify the game to their own advantage and not necessarily to the socially optimal outcome. Players realize that their promised payments to other players influence the strategic properties of the game, and they each try to set side contracts to optimally manipulate the overall play of the game. In some limited contexts, it turns out that players' incentives to manipulate the play in the game when aggregated across players actually coincides with the social optimum. But in many other cases the private side contracting incentives and the social welfare will remain at odds. Moreover, we show that there are some interesting strategic differences in how such side contracting affects games with two players versus games with more than two players.

In the remainder of the introduction we discuss the contribution of this work to the literature. In Section 2 we discuss an the public goods contribution game at length to illustrate some of the points of the paper and to show why the simple Coasian intuition might fail. In Section 3 we present the formal model. Results for the two player case appear in Section 4, and then in Section 5 we present some applications of the results. In Section 6 we present the analysis for three or more players. Finally our concluding remarks appear in Section 7.

Contributions to the Literature

As is clear from the discussion above, this paper is related to what has become known as the Coase theorem. Coase (1960) was not explicit about the type of agreements between agents that are necessary as a form of bargaining to reach efficiency. The types of contracts that we explore here are fairly rich in that they may be contingent on any subsequent actions, and certainly allow players to write contracts that would lead to efficiency. It is simply not in their strategic interests to do so. Moreover, as we shall discuss below, the analysis is fairly robust to the timing of side contracting.

While our results might be viewed are somewhat negative, there are various interpretations that one might have. A cautious interpretation of our results is that we show that the way in which side contracts are proposed (and accepted) is critical to determining whether or not efficiency will be reached, even with full information and the absence of any friction. The issue is more subtle than one might imagine, and if efficiency is to be supported it will necessarily require some fairly complicated side contracting. We return to discuss this in more detail in the concluding remarks, as after we have presented the results we can be more specific about how the structure of contracting matters.

Another important point to keep in mind is that when our results imply that such side contracting fails to result in efficient outcomes, we do not have to think of this as a refutation of efficiency per se. Rather, it suggests that we should then see the emergence of other institutional or exogenous solutions. That is, when the necessary conditions for supporting efficiency fail, mechanism design has a significant descriptive as well as prescriptive role in economics. We should expect to see institutions emerge such as: restrictions on the class of admissible side contracts, the use of bonding agents, and mutually binding third party adjudication precisely for this class of situations. On the other hand, when side contracting implies efficiency, then mechanism design becomes secondary — the players can make binding contracts to hit upon the efficient outcome themselves.

Our analysis also relates to some other work in contracting theory. In particular, the reason inefficiency arises in our setting is different from that in the rest of the literature. As mentioned above, when inefficiency arises in the setting we consider, it comes from the fact that each agent is attempting to offer transfers that manipulate other agents' behavior to his or her advantage, and not necessarily to what is socially desirable. Agents effectively compete for payoffs and this can lead to behavior that is ultimately inefficient. We emphasize that this is a different contracting failure from the ones that have been the primary focus of the recent contracting literature.[1] Here, there is no asymmetry of information, no cost to contracting, no unforeseen contingencies or complexity to contracting, and contracts are fully binding and enforced. The failure to reach efficiency arises from the combined in-

[1]Much of the recent contracting literature has focused on imperfections related to costs of contracting, asymmetric information, limited enforcement of contracts, and non-verifiability of actions or information. For a recent overview of this extensive literature, see MacLeod (2001). Anderlini and Felli (2001) provide a nice discussion of the relationship of that literature to failures of the Coase theorem.

centives of each party to manipulate the behavior of the others through the contracts. Thus, the work here points out that one cannot abstract away from the process by which contracts are chosen, as that process itself can affect whether efficiency is obtained.

A virtue of our approach is that much of the existing contracting literature can be embedded in our framework as special cases with added restrictions on the admissible contracts. Here we allow for a very general class of side payments, especially in terms of the contingencies possible. In particular, a side payment rule is function from the space of joint strategies to nonnegative numbers. This is a richer payment class than is admitted in much of the contracting literature. Thus, our results provide a robustness check on these papers, and perhaps a rationale for the emergence of contracting restrictions. For example the common agency literature is a special case of our model where only the players labeled "principal" are allowed to make offers, and the admissible transfers can only depend on the actions of the agents. Those limitations, can result in different predictions (e.g., see Prat and Rustichini (1999)).[2]

Our results can also be viewed as providing an important warning to the vast literature on mechanism design and implementation. With very few exceptions, the game theory, mechanism design, and implementation literatures analyze games or mechanisms when they are taken as fixed by the players. The strategies of the game or mechanism are assumed to completely capture those available and the payoffs capture all of those relevant. Hurwicz (1994) offers compelling arguments for viewing mechanisms in a larger natural context. He discusses a variety of factors, ranging from enforceability of the outcomes, to natural actions that are available to agents outside of those of the mechanism. Hurwicz argues that such factors may ultimately play important roles in determining the outcome of the mechanism, and so

[2]Another application is the contracting externalities literature. For example in Aghion and Bolton (1987) there are three players, the incumbent seller, a customer and an entrant. They show that the customer and incumbent may contract to an inefficient outcome that deters entry. In their framework the entrant is not allowed to make pre-game offers to the incumbent or the customer. Segal (1999) shows how many contracting papers can be unified by the concept of a contracting externality. Again our results provide insight into the role played by the restrictions on the class of contracts used in these papers. A further example and illustration of some of these points can be found in the public goods literature discussed at length below.

these must be taken into account in designing mechanisms, and even more basically the effects of such factors must be understood when making predictions concerning the performance of any given mechanism. One can view our analysis from this perspective, taking the ability of agents to commit to transfers prior to a game as something that is natural and unavoidable in many contexts. Our findings show that the outcomes of the game can be greatly affected by thinking of games and mechanisms in a larger context, which lends support to Hurwicz's point. Most importantly, the implications can go in different directions. As mentioned above and shown below, in some instances the broader endogenous view of games leads to improvements in efficiency and other times it actually worsens the outcomes. This emphasizes how important it is to understand the broader context in which a game or mechanism resides, and suggests this as an important area for further study.

Of course, with respect to the game theoretic literature, there is an easy way to view our work. We characterize the outcomes of games when strategy contingent side payments are allowed to be promised by players.

One exception to viewing a game as fixed is delegation games (e.g., see Fershtman, Judd, and Kalai (1991) and a recent application in Miller and Pazgal (2001)).[3] In delegation games players may hire another player to play the game for them. This effectively allows a player to change their own incentives and thus can change the outcomes of a game. There are important differences between the delegation setting and the side contracting setting we consider. Most importantly in delegation games a player can only change their own payoff structure, and cannot make promised payments to other players to induce other players to change their strategies. Also, in our setting players' own payoffs can only be changed via transfers to or from another player, rather than to a delegate. Thus there is also a difference in the way payoffs in the game must be balanced. These differences prove critical and lead to completely different results. Nevertheless, while the delegation setting is quite different from what we analyze here, there are similarities in the underlying motivation and both can be viewed as forms of endogenous

[3] A few other exceptions are the analysis of choices of mechanisms by competing sellers (e.g., see McAfee (1993)), choice of voting rules (e.g., see Barbera and Jackson (2000)), flexibility on the part of the planner (e.g., Baliga and Sjöström (1995), and mechanism selection more generally (e.g., see Lagunoff (1992) and Allen (2001)). For recent introductions to the implementation and mechanism design literatures, and some additional discussion of endogenous mechanisms, see Jackson (2000, 2001).

games.

Another exception to viewing the game as fixed is found in Kalai and Samet (1985) who looked at players trying to come to a unanimous and binding agreement as to a social state that is to be chosen.[4] They find that given enough rounds in the unanimous voting process a Pareto efficient agreement about the social state will be reached. The social states can include a specification of how a second-stage game will be played, and so our work is related to the Kalai and Samet analysis. However, their analysis focuses on the unanimity voting process over social states and then assumes the social state to be binding — so if one takes the social state to be play of a game, then it is implicit that players can commit to how they will play the second-stage game. In our analysis players only commit to contingent transfer functions, and then the play in the second-stage game must be an equilibrium given the transfer functions. So we analyze what transfers might be made to support play in the second-stage game, and see when it is in the interest of players to make the transfers that actually support efficiency as opposed to making transfers to otherwise manipulate the game.[5] In the Kalai and Samet analysis players come to agree on an efficient outcome, while we will find that explicit modeling of transfers among players will change the analysis and sometimes lead to inefficiency, at least for two players, while the results will be more congruent for three or more players.

Finally, our analysis is closely related to the study of matching games that have been analyzed by Guttman (1978, 1987), Danziger and Schnytzer (1991), Guttman and Schnytzer (1992), and Varian (1994a), among others. They show that efficiency can be obtained when agents can undertake matching plans in the context of public goods and some other settings with externalities. Our results will in many contexts be at odds with the results from those papers. The reason behind the difference in results is that those papers limit the set of contracts that are available to agents so that they can only make some types of promised payments. As we will point out, if agents

[4]See also Kalai (1981), as well as later work by Bensaid and Gary-Bobo (1996) who consider a single player who suggests contracts on how to play a game.

[5]One could also allow Kalai and Samet's social states to include specifications of transfer functions and how the game will be played. However, their efficiency result only holds for finite sets of social states, and so the richness of transfer functions that is critical in our analysis could not be admitted in their setting. Indeed, we will reach a different conclusion regarding efficiency which hinges on this richness.

can choose from a richer set of contracts (and they have a strict incentive to, as we shall see) then efficiency no longer holds. This is illustrated in the next section and provides a starting point for our analysis.

2 An Example: Public Goods and Matching Contributions

The intuitive argument for how side contracting might support efficient outcomes, and the reasoning of Coase, is roughly as follows. One agent can offer a second agent compensation as a function of the second agent's action that effectively reflects any externality that the second agent's action has on the first agent. Essentially, it is as if the first agent says to the second agent "the benefit to me is x if you take action A rather than action B, and the cost to you of taking action A rather than B is only y where $x > y$, and so I will pay you z, where $x \geq z \geq y$, if you play A instead of B." Any z such that $x \geq z \geq y$ will provide sufficient incentives.

The more subtle issue arises when this is put into a richer setting, where more than one agent is taking an action at the same time. Then strategic factors come into play that make the analysis significantly more complicated.

To make things concrete, let us consider an example which has been well-studied in the literature. The example is of a voluntary provision of a public good. Individuals may each make a contribution to cover the cost of the provision of a public good, and the amount of the good provided is determined by the total of the donations. As is well-known, in such contribution games a free rider problem exists and equilibrium behavior results in inefficiently low contributions.[6] When deciding on a contribution, each individual considers the marginal impact that his or her contribution makes to his or her own well-being, but not to the social impact it has on others who are also consuming the public good. However, such a game should not be viewed in isolation. If the economic agents understand the incentives and inefficiency of the outcome of voluntary contributions game, they can work beforehand

[6] For some recent work that examines the possibility of reaching efficiency through sequential contributions, see Fershtman and Nitzan (1991), Admati and Perry (1991) and Marx and Matthews (2000). For some approaches to (re-)designing games to lead to efficiency see Bagnoli and Lipman (1989), Jackson and Moulin (1992), Varian (1994b), Ellis and van den Nouweland (1998), Bag and Winter (1998).

to correct it. In a Coasian manner, they can side contract to reward each other for making contributions, and try to correct each other's incentives. For instance, an individual can announce that they will "match" (at some rate) the contributions that are made in the game.[7]

For example, consider a society of two individuals and a public good produced through a linear cost technology. For simplicity, consider a case where each individual can either contribute a unit of the public good or not. The utility of one unit of the public good is 4 to each agent. The marginal utility of a second unit of the public good is 3. To contribute to the public good costs an agent 5 units of utility. So, if only one agent contributes to the public good then that agent's total utility is 4-5=-1 and the other agent's utility is 4. If both agents contribute then they each earn 4+3-5=2. The payoffs are represented as follows.

	C	N
C	$2, 2$	$-1, 4$
N	$4, -1$	$0, 0$

This has the classic form of a prisoner's dilemma, with a unique equilibrium involving no contribution of the public good.

Now let us examine the Coasian intuition and what happens if agents can make binding offers of action contingent side payments before the game is played. Consider the efficient situation where both players are contributing to the public good. The column player would gain 2 by deviating and not contributing. This deviation would hurt the row player by 3. So it is in the row player's interest to offer any payment of at least 2 and no more than 3 to the column player contingent on the column player contributing. The only such payment that makes sense from the row player's perspective is a payment of 2, since giving any more is simply a gratuitous transfer to the other player. The same logic works in reverse, so that the column player is willing to make a payment of 2 to the row player contingent on the row

[7] In this paper we model direct side payments rather than actions such as matching contributions. Whether a player is partly reimbursed for a contribution, or someone else makes a matching contribution is irrelevant to the outcome. Moreover, the direct side payments allow for a more general set of contingent payments that go beyond simple matchings; and such side payments can be made in any game moving beyond public goods contributions games.

player contributing. Taking these two transfers into account, the net payoffs to the two players looks as follows.

	C	N
C	$2,2$	$1,2$
N	$2,1$	$0,0$

This side contracting has changed the game so that the efficient contributions are an equilibrium (and in this example are actually weakly dominant strategies). This insight is first due to Guttman (1978), and has been extended to a variety of voluntary contribution games and other games with externalities by Danziger and Schnytzer (1990), and Varian (1994), among others.

This, however, is not the end of the story. We can ask whether these particular side payments are in fact part of equilibrium play. For instance, if the column player is offering the row player a payment of 2 contingent on the row player contributing, is it in the row player's interest to offer a payment of 2 contingent on the column player contributing? The answer is no.[8] Suppose that instead, the row player offers to pay the column player $1+\varepsilon$ for each unit of contribution to the public good, made by either player (where $\varepsilon > 0$).[9] The resulting game is as follows.

	C	N
C	$2-2\varepsilon, 2+2\varepsilon$	$-\varepsilon, 3+\varepsilon$
N	$3-\varepsilon, \varepsilon$	$0,0$

This has a unique equilibrium which is inefficient, but better from the row player's perspective: the column player contributes and the row player does

[8]The reason that this does not contradict the analyses of Guttman (1978), Danziger and Schnytzer (1990), and Varian (1994a), is that they only consider a limited form of matching contracts, where matching or payments can only be made in proportion to the actions taken by the other agents (see also Qin who considers also payments made only contingent on own action). As we see here, either player would strictly gain by deviating and using a different sort of contract (arguably just as simple). In fact one observes matching offers of the form "I will match any contributions," rather than just "I will match the contributions made by other people." Such seemingly minor differences in contract specification have important implications for incentives, as illustrated in this example.

[9]If $\varepsilon = 0$ there are two equilibria to the game, but both are still inefficient and involve the row player not contributing.

not contribute. Thus, it will not be an equilibrium for the players to offer the "efficient" matching plans as this example shows. As will follow from the theorems we prove below, the only equilibria to this two stage process for this example are inefficient — (and in fact in this example are in mixed strategies).

Let us now provide a general framework for analyzing these issues.

3 Definitions

A set $N = \{1, \ldots, n\}$ of players interact in two stages.

First, let us offer an informal description of the process.

Stage 1: Players simultaneously announce transfer functions. That is, each player announces a profile of functions indicating the payments that they promise to make to each other player as a function of the full profile of actions chosen in the second-stage game.

Stage 2: Players choose actions in the game.

Payoffs: The payoff that player i receives is his or her payoff in the game plus all transfers that other players have promised to i conditional on the actions played in the game minus the transfers player i promised to make to other players conditional on the actions played in the game.

The transfer functions that are announced in Stage 1 are binding. There are many ways in which this could be enforced, ranging from reputation, posting a bond with a third party, to having legal enforcement of contracts. We do not model these details here, and simply assume that the announcements are public so that all players know all of the promises. As will become clear in the analysis that follows, the simultaneity of the announcements is not critical to the overall theme of the results (that efficiency is only reached in limited situations), but is important in determining the relative payoffs of the players.

The fact that the transfer promises are unilaterally binding fits well with the matching contributions in public goods games, reward schemes and a variety of other applications.[10]

[10]For an alternative framework where unanimity is required to enforce another's offer see Ray and Vohra (1997) or Qin (2001).

We also point out that players can effectively refuse a contract by simply announcing a transfer that undoes the other player's transfer.[11] In fact, in this way they can choose to undo part of the other player's promise and keep part.

Let us also say a word or two about "consideration." Some contracts that we admit are of the form where one player makes a promise contingent only on his or her own action. These contracts are useful ways of posting a bond or committing ones-self to taking certain actions. Such contracts might not always be legally enforceable because of the lack of "adequate consideration" by the other player - i.e., the other player did not do anything. Note however, that these promises are easily approximated by promises that vary in some way on the other players' actions, or can be supported by other considerations outside of our model such as reputation, as we have already discussed.

We now provide formal definitions.

The Underlying (Second-Stage) Game

The second-stage game consists of a finite pure strategy space X_i, with $X = \times_i X_i$. Let $\Delta(X_i)$ denote the set of mixed strategies for player i, and let and $\Delta = \times_i \Delta(X_i)$. We denote by x_i, x, μ_i and μ generic elements of X_i, X, $\Delta(X_i)$, and Δ, respectively. In some cases we use x_i and x to denote elements in $\Delta(X_i)$ and Δ, respectively, that place probability one on x_i and x.

The restriction to finite strategy spaces provides for a simple presentation of the results, avoiding some technical details. Nevertheless, games with a continuum of actions are important, and we provide results for the case of games with continuous action spaces in the appendix. These results are a straightforward extension of the finite case.

Payoffs in the second-stage game are given by a von Neumann-Morgenstern utility function $v_i : X \to I\!R$.

The First Stage Transfer Functions

The transfer functions that player i announces in the first period are given the vector of functions $t_i = (t_{i1}, \dots, t_{in})$, where $t_{ij} : X \to I\!R_+$ represents the

[11] Note also, that this can be added to a transfer that the second player wants to announce. This is not possible in the context of mixed strategies, but our focus is on support in pure strategies where this is possible.

promises to player j as a function of the actions that are played in the second-period game. So, if x is played in the second period, then i transfers $t_{ij}(x)$ to player j.

Let $t = (t_1, \ldots, t_n)$. Also, denote by t_i^0 the degenerate transfers such that $t_{ij}^0(x) = 0$ for all $x \in X$, and let $t^0 = (t_1^0, \ldots, t_n^0)$.

The Payoffs

The payoff to player i given a profile of transfer functions t and a play x in the second-period game is then[12]

$$U_i(x,t) = v_i(x) + \sum_{j \neq i} (t_{ji}(x) - t_{ij}(x)) \,.$$

So, given a profile of transfer functions t and a mixed strategy μ played in the second-period game, the expected utility to player i is

$$EU_i(\mu,t) = \sum_x \times_i \mu_i(x_i) \left[v_i(x) + \sum_{j \neq i} (t_{ji}(x) - t_{ij}(x)) \right].$$

Let $NE(t)$ denote the set of (pure and mixed) Nash equilibria of the second-stage game where payoffs are given as above. So this is the set of Nash equilibria taking a profile of transfer functions t as given, and only varying the strategies in the second-period game.

Supportable Strategies and Payoffs

A pure strategy profile $x \in X$ of the second-stage game together with a vector of payoffs $\overline{u} \in I\!R^n$ such that $\sum_i \overline{u}_i = \sum_i v_i(x)$ is *supportable* if there exists a subgame perfect equilibrium of the two stage game where some t is played in the first stage and x is played in the second stage (on the equilibrium path), and $U_i(x,t) = \overline{u}_i$.

Supportability is a condition that applies to a combination of a strategy profile and a set of payoffs. We refer to both since in some cases transfers must be made on the equilibrium path to support x as part of an equilibrium. In such cases the payoffs including transfers differ from the original underlying payoffs without transfers.

[12]This assumes transferable utility, and it would be interesting to see how this extends to situations where private goods transfer at different rates across players.

The definition supportability looks at pure strategies in terms of what is played on the equilibrium path. In many games (in fact generically), there is a unique x that is efficient. Thus, it makes sense to focus on pure strategy equilibria, at least in terms of the second period. The focus on pure strategies in terms of t's is for technical convenience, as the space of mixed strategies over all such transfer functions is a complicated animal (measures over functions) and considering them would not add much to the results.[13]

Surviving Equilibria

In addition to understanding supportability in the two stage process, we are also interested in following question. When does an equilibrium of the original underlying game survive to be supportable when the two stage process is considered?

Consider a pure strategy profile $x \in X$ of the second-stage game that is an equilibrium of the second stage when no transfers are possible ($x \in NE(t^0)$). Such an equilibrium *survives* if there exists a subgame perfect equilibrium of the two stage game where some t is played in the first stage and x is played in the second stage (on the equilibrium path), with net payoffs being $U_i(x,t) = v_i(x)$.

Note that x survives if and only if x is a Nash equilibrium of the second-stage game and $(x, v(x))$ is supportable in the two stage process. Together, the notions of supportability and surviving then give us an idea of how the set of equilibrium and equilibrium payoffs change when players can make binding transfer commitments.

4 Two Player Games

The results for two player games and games with more than two players differ significantly and so we treat them separately. We start with an analysis of two player games.

The following notion plays an important role in the characterization results that follow.

[13]In many of the examples where efficiency turns out not to be supportable in pure strategies, allowing for mixed strategies would not help. We are not sure whether this is always the case. We do know that it is important to consider mixed strategies off the equilibrium path in the second-period game, and so we do explicitly account for this.

Solo Transfers

Suppose that only one player were allowed to propose transfers in the first stage. We can consider the transfers that would be best from this player's perspective.

Let[14]

$$u_i^s = \sup_{t_i}\left[\min_{\mu \in NE(t^0_{-i},t_i)} EU_i(\mu, t^0_{-i}, t_i)\right].$$

So, a player's "solo" payoff is the one obtained when the player is allowed to announce any transfer function that he or she likes and no other player can make any transfers. As there may be several equilibria in the second-stage game that result from any given transfer function, we must have some idea of which one will be relevant. This definition imagines the worst continuation equilibrium for i once t is in place. This turns out to be the correct definition for characterizations that follow.

To get some intuition for the definition above, consider the public goods game from Section 2:

	C	N
C	$2,2$	$-1,4$
N	$4,-1$	$0,0$

Here the solo payoff for the row player (and similarly the column player) is 3. By making a payment of $1+\varepsilon$ conditional on the column player playing C, the new matrix becomes

	C	N
C	$1-\varepsilon, 3+\varepsilon$	$-1,4$
N	$3-\varepsilon, \varepsilon$	$0,0$

This has a unique equilibrium leading to a payoff of $3-\varepsilon$ for the row player. Taking the sup over such payments leads to a payoff of 3 for the row player, which is the solo payoff. The usefulness of this concept is illustrated in the theorems below.

[14]Note that the min in this expression are well-defined since the set of Nash equilibria of a finite game is compact. The sup is necessary as there may be no maximizer. For instance, consider a game where $X_1 = \{x_1\}$ while $X_2 = \{x_2, x_2'\}$. Let player 2 be indifferent between the actions and player 1 prefer that 2 play x_2. Any positive transfer from 1 to 2 leads to a unique equilibrium of x_2, but a 0 transfer leads to a minimizing equilibrium of x_2'.

Our first result is a characterization of the Nash equilibria of a game that survive when transfers are introduced.

THEOREM **1** *If $n = 2$, then a Nash equilibrium x of the underlying game survives if and only if $v_i(x) \geq u_i^s$ for each i. Moreover, if x survives then there is an equilibrium in the overall process where no transfers are made in the first stage and x is played in the second stage.*

The formal proof of Theorem 1 uses the proof of Theorem 3 and appears in the appendix. However, the intuition is fairly simple and we explain it here.

First, let us show that this condition is sufficient to have x survive. Consider a Nash equilibrium x such that $v_i(x) \geq u_i^s$. On the equilibrium path let players make no transfers in the first stage (play t^0) and then play x in the second stage. Off the equilibrium path, if some player offers transfers in the first period, then identify the worst equilibrium for that player in the resulting second-stage game and have that be played in the continuation (and if more than one player offers transfers in the first period then play any equilibrium in the second stage). This is easily seen to be a subgame perfect equilibrium of the overall game: the best payoff a player can get by deviating in the first stage is no more than their solo payoff, which is not improving; and given that no transfers are made in the first stage, x will be an equilibrium in the second stage.

Next, consider a Nash equilibrium x that survives. We argue that $v_i(x) \geq u_i^s$. Let t be any transfer function that is made in the first stage as part of an equilibrium where x is played in the second stage. Suppose to the contrary that $v_i(x) < u_i^s$. The definition of solo payoff then implies that some transfer $\bar{t}_i$, so that the payoff to i under worst continuation equilibrium under $\bar{t}_i, t_j^0$ is higher than $v_i(x)$. So, let i do the following: make a transfer that cancels out the transfers under t_j and then adds $\bar{t}_i$ on top. That is, let i announce $\widehat{t}_i = \bar{t}_i + t_j$. Note that the pair of transfers $\widehat{t}_i, t_j$ leads to exactly the same second stage payoffs as $\bar{t}_i, t_j^0$. Thus, from the definition of $\bar{t}_i$, it follows that if i deviates to $\widehat{t}_i$ while j plays t_j then even the worst continuation equilibrium in the second stage will result in a payoff which is higher than $v_i(x)$. This contradicts the fact that t was part of an equilibrium where x was played in the second stage.

To get an idea of the impact of Theorem 1, we illustrate it in the context of several examples.

Example **1** *Only the Efficient Equilibrium Survives.*

	A	B
A	$2, 2$	$0, 0$
B	$0, 0$	$1, 1$

In this pure coordination game, there are two equilibria (A,A) and (B,B). The solo payoffs are 2 for each player[15] and so it follows from Theorem 1 that the only equilibrium that survives once transfers are allowed is (A,A).

Example **2** *The Efficient Equilibrium Does Not Survive.*

Consider the following game, which has an efficient equilibrium of U,L leading to payoffs of 2,2. It is easily checked that the solo payoffs are 3 to each player.

	L	C	R
U	$2, 2$	$0, 0$	$0, 0$
M	$0, 0$	$3, 0$	$0, 0$
D	$0, 0$	$0, 0$	$0, 3$

So, in this game the efficient equilibrium does not survive. In fact, no equilibrium survives, and the only equilibria of the two stage process involve mixing over transfer functions announced in the first stage.

Next, we provide an example which shows why it was necessary to consider mixed strategies in the definition of solo payoffs.

Example **3** *Mixed Strategies in the solo Definition*

Consider the following game.

	L	C	R
U	$1, 10$	$0, 0$	$0, 0$
M	$0, 0$	$3, 0$	$0, 10$
D	$0, 0$	$0, 10$	$3, 0$

[15]For instance, if the row player makes the transfer of 2 to the column player conditional on (B,A) being played, then the unique equilibrium becomes (A,A).

Let us see if the strategies $x = (U, L)$ with payoffs $\overline{u} = (1, 10)$ are supportable. (Here the minimal transfer function is t^0.) The highest payoff in the matrix for the column player is 10, and so this should be fine. So, we need only consider what the row player can expect. If we use the straight pure strategy solo definition, then the best payoff that the row player can induce is 1. This would suggest that U,L would in fact be supportable without any transfers. However, if we consider mixed strategies things change. Suppose that the row player pays the column player 2 conditional on U being played. That leads to the following matrix.

	L	C	R
U	$-1, 11$	$-1, 1$	$-1, 1$
M	$0, 0$	$3, 0$	$0, 10$
D	$0, 0$	$0, 10$	$3, 0$

This game has a unique Nash equilibrium which is a mixed strategy equilibrium (equal mixing on M and D by row, and C and R by column) leading to a payoff of 1.5 to the row player. This means that the threat point for the row player is indeed 1.5. That is what is captured in the definition of solo payoffs.

As one can also verify that the modified solo payoff of the column player is 10, and so Theorem 3 below will actually show that no action-payoff pair is supportable in this example. What that means is that any equilibrium in the two stage process will involve mixing over the transfers functions announced in the first stage.[16]

Note that the reasoning behind Theorem 1 did not rely on the fact that x was a Nash equilibrium of the second-stage game to begin with. So, in fact we have just argued a necessary condition for supportability as well as survivorship. This is stated in the following Theorem:

Theorem 2 *If $n = 2$, then $(x, \overline{u})$ is supportable only if $\overline{u} \geq u_i^s$ for each i.*

While there are instances where $\overline{u} \geq u_i^s$ for each i is enough to ensure that $(x, \overline{u})$ is supportable, there are other games where it is not. The full necessary and sufficient condition for supportability is defined below.

[16] Since the total of the modified solo payoffs is 11.5, any pure strategies of transfers in the first stage must lead to an equilibrium in the second stage that is below the modified solo payoff of one of the two players who could then benefit by deviating.

Before stating the full characterization, we note that Theorem 2 can still be quite useful. This follows since $\overline{u} \geq u_i^s$ is already a very demanding condition that fails in many games. Thus, supportability can be ruled out in many cases simply by checking whether $\overline{u} \geq u_i^s$.

That is true of the public goods contribution example in Section 2, as there the solo payoff of each player is 3, and there is no pair of actions possible that could lead to a payoff of at least 3 to each player at the same time.

Now let us turn to obtaining the full characterization of supportability.

Consider supporting $(x, \overline{u})$. It is important to allow $\overline{u} \neq v(x)$, as x may not naturally be a Nash equilibrium to begin with and so some side payments may be necessary to support an action profile as an equilibrium. Thus, in order to characterize supportability it will be important to have some idea of what transfers are (minimally) necessary.

In fact, it is easy to define the minimum necessary set of transfers to support a given action profile as an equilibrium. First, if $\overline{u}_i \neq v_i(x)$ for some i, then it must be that some transfer is being made. In particular, if $v_i(x) > \overline{u}_i$ the it must be that i is transferring at least $v_i(x) - \overline{u}_i$ to j. Similarly, if $v_i(x) < \overline{u}_i$ the it must be that j is transferring at least $\overline{u}_i - v_i(x)$ to i. Also, to preclude other deviations if $v_i(x_j, \widehat{x}_i) > \overline{u}_i$, then it must be that if $x, \overline{u}$ is supported then i must be making a transfer $t_{ij}(x_j, \widehat{x}_i) \geq v_i(x_j, \widehat{x}_i) - \overline{u}_i$, otherwise x would not be an equilibrium in the second stage of the process. Collecting these ideas leads to the following definition.

Minimal Transfers

The *minimum transfer function profile* $t^{x,\overline{u}}$ for a pair $x, \overline{u}$ is defined by:

$$t_{ij}^{x,\overline{u}}(\widehat{x}) = \begin{cases} \max[v_i(\widehat{x}) - \overline{u}_i, 0] & \text{if } \widehat{x}_j = x_j \\ 0 & \text{otherwise.} \end{cases}$$

The idea of a minimum transfer function is illustrated in the following example.

Example 4 *Minimal Transfer Function*

Consider the following game.

	L	R
U	$4,4$	$0,6$
D	$5,0$	$0,6$

Let us consider supporting the actions (U,L) as part of an equilibrium. No matter what transfers are promised, the column player can always get at least a payoff of 6 by not announcing any transfers in the first stage and then playing R in the second stage. So in order to support the efficient combination of (U,L) as part of an equilibrium, it must be that a transfer of at least 2 is made from the row player to the column player conditional on U,L being played. This gives us one part of the minimal transfer function. So, now consider supporting $x =$(U,L) with payoffs $\overline{u} = (2, 6)$. In order to have (U,L) be an equilibrium with these payoffs, it would also have to be that the row player transfers at least 3 to the column player conditional on (D,L), as otherwise (U,L) could not be an equilibrium (given the transfer of 2 from the row player to the column player conditional on (U,L)). So, these transfers are the minimal transfers to support $x =$(U,L) and $\overline{u} = (2, 6)$.

To see why these must be accounted for in the definition of solo payoff in order to get a full characterization, modify the payoff of (D,L) to be (5,4). In that case, the minimal transfers lead to a new payoff of (D,L) of (2,7). That makes the column player's modified solo payoff 7 (as defined below) while the column player's solo payoff was only 6. This higher payoff is the relevant one for strategic considerations in the two stage game where one tries to support (U,L) as an equilibrium.

Given the definition of minimum transfers above, we can now define the solo payoffs noting that these minimal transfers (or some larger transfers) would have to be in place in order to lead to $x, \overline{u}$ as part of an equilibrium outcome in the two stage process.

Modified solo Payoffs

$$u_i^{ms}(x, \overline{u}) = \sup_{t_i} \left[\min_{\mu \in NE(t_{-i}^{x,\overline{u}}, t_i)} EU_i(\mu, t_{-i}^{x,\overline{u}}, t_i) \right].$$

The above definition of modified solo payoffs leads to the following complete characterization of supportable action-payoff pairs.

Theorem **3** *If $n = 2$, then $(x, \overline{u})$ is supportable if and only if $\overline{u}_i \geq u_i^{ms}(x, \overline{u})$ for each i. Moreover, if $(x, \overline{u})$ is supportable it is supportable with the minimal transfer function profile $t^{x,\overline{u}}$.*

5 Applications

To see the implications of the results above, let us examine some common settings.[17]

One Sided Externalities

Consider a classic one sided externality, such as Coase's example of a steel mill affecting a laundry. Let x_1 denote the output of the steel mill and x_2 denote the output of the laundry. The utility functions are given by $v_1(x_1)$ and $v_2(x_1, x_2)$. Let there be a unique Nash equilibrium x_1^n, x_2^n, and a unique efficient point x_1^*, x_2^*.

Let us consider supporting the efficient solution. Player 2's minimal transfer function is $t_{21}(x) = v_1(x_1^*) - v_1(x_1^n)$ if $x_1 = x_1^*$ and $t_{21}(x) = 0$ otherwise.[18] Let $\overline{u}$ be the induced utility levels under efficiency with t_{21} as specified above, and t_{12}^0.

Player 1 gets $\overline{u}_1 = v_1(x_1^n)$ under any Nash equilibrium with t_{21}, t_{12}^0. By offering any other t_{12}, player 1 will make additional payments without gaining any additional utility. Thus, $u_1^{ms}(x^*, \overline{u}) = \overline{u}_1$. Given t_{12}^0, $u_2^{ms}(x^*, \overline{u})$ is the same as player 2's solo payoff u_2^s. That payoff is the solution to

$$\max_{x_1, x_2, t_1} v_2(x_1, x_2) - t_1$$

subject to

$$v_1(x_1) + t_1 \geq v_1(x_1^n).$$

It is clear that a solution must involve setting $t_1 = v_1(x_1^n) - v_1(x_1)$ and so we can rewrite the problem as

$$\max_{x_1, x_2} v_2(x_1, x_2) + v_1(x_1) - v_1(x_1^n).$$

[17] The applications we consider are standard ones, and as such generally have some feature of strategic substitutes or complements. However, as seen from the characterization theorem, the conditions for supportability are not related to conditions of strategic complementarity (or sub- or super- modularity), but instead depend on the modified solo payoffs. This can be seen by looking across examples, or even simply at the prisoners' dilemma example.

[18] This transfer function is discontinuous, but one can also work with $t_{21}(x) = v_1(x_1) - v_1(x_1^n)$, which is continuous. For details on the treatment of continuum actions, see the appendix.

The solution to this is the efficient production x^*, and involves making the payment $t_{21}(x^*) = v_1(x_1^n) - v_1(x_1^*)$. Thus, by Theorem 3, the efficient point is sustainable as a contracting equilibrium. Coase's intuition that the polluter and victim can reach an efficient outcome is indeed correct with these side contracts for a one sided externality.

Bertrand Competition

Consider the case of two firms competing in a Bertrand market. Let each firm have a linear cost function $c(q_i) = cq_i$ as a function of their production quantity q_i, and the demand function be described by $Q(p)$ where $Q = \sum q_i$ and p is the lowest price offered by any firm. Here the strategic variable of each firm is price $p_i \in I\!R_+$. Follow the textbook Bertrand rule that firms charging the lowest price split the market evenly and that firms with higher prices sell zero. The Nash Equilibrium payoff for each firm in the underlying Bertrand game is zero. Let π^m denote the industry payoff if all firms charge the monopoly price. Can we support the strategy $p_i = p^m$ and payoffs $\overline{u}_i = \frac{\pi^m}{2}$ for all $i \in N$?

Consider supporting these with the corresponding minimal transfer functions $\overline{t}$, which in this case amount to paying $\frac{\pi^m}{2}$ to the other firm if the other firm j charges the monopoly price and firm i does not and 0 otherwise.[19] Note that a firm can get arbitrarily close to the full monopoly price by slightly lowering price, which would increase their profits by $\frac{\pi^m}{2}$. So these must be the minimal supporting transfer functions. Let us then consider the modified solo payoff to firm 1, u_1^{ms}. This modified solo payoff is exactly $\frac{\pi^m}{2}$. It is straightforward to check that the best player 1 can hope to get is for the players to both charge the monopoly price and split the market — which requires paying player 2 half of the profit to keep him from lowering his price slightly and getting arbitrarily close to the full profit. Thus the payment to player 2 cancels that from 2 to 1. Therefore, by Theorem 3 the efficient (collusive) outcome is supportable.

Public Goods

Consider the two person example of voluntary contributions to a public good. Let $x_i \in I\!R_+$ be player $i's$ contribution and her utility be $v_i(x_1, x_2) = 2\theta_i(\sum x_j)^{\frac{1}{2}} - x_i$ where $\sum \theta_j = 1$ and $\theta_j > 0$ for $j = 1, 2$. Suppose that

[19] These are more than the minimal transfer functions in some cases where i undercuts at a price that leads to less than monopoly profits, but these transfers are simple and suffice in this problem.

$\theta_1 > \theta_2$. This ensures a unique Nash equilibrium in the contribution game, $x_1^n = \theta_1^2, x_2^n = 0$. The associated utilities are $v_1(x_1^n, x_2^n) = \theta_1^2$, $v_2(x_1^n, x_2^n) = 2\theta_2\theta_1$. The efficient allocation is any pair such that $\sum x_i = 1$. Moreover the net utilities at any efficient allocation sum to 1, that is if the pair $(x', \overline{u})$ is efficient and sustainable then $\sum \overline{u}_i = 1$.

Now consider 1's solo payoff u_1^s. Let $t_{12}(x_2) = \theta_1 x_2$. If this offer is made then in the subgame perfect equilibrium that follows $x_2 = 1$. Thus, $u_1^s \geq 2\theta_1 - \theta_1 = \theta_1$.

Now consider 2's solo payoff u_2^s. Player 2, knowing that absent any offer she can free ride on 1's Nash contribution, can use the minimal transfer to force player 1 to provide the efficient level of contribution. In particular set $t_{21}(x) = \theta_1^2 - (2\theta_1 - 1)$ if $x_1 = 1$ and $t_{21}(x) = 0$ otherwise.[20] If this offer is made then in the subgame perfect equilibrium $x_1 = 1$. Thus, $u_2^s = 2\theta_2 - [\theta_1^2 - (2\theta_1 - 1)] = 1 - \theta_1^2$.

Putting these together we find that $\sum u_i^s \geq \theta_1 + (1 - \theta_1^2) > 1$. Thus as at any efficient outcome $\sum u_i(x', t') = 1$ it follows from Theorem 2 no efficient outcome is supportable. Indeed, there is no pure strategy subgame perfect equilibrium of the two-stage game.

Agency Models

Another application of our results is to the agency problem. A simple principal-agent setting may be viewed as a game with one sided externalities, where only the agent takes actions in the second-stage game, and so the principal only worries about making transfers to the agent to induce efficient actions. Since this is a one sided externality problem, then as noted above efficient outcomes can be supported.[21]

A more interesting case is that of common agency (e.g., see Bernheim and Whinston (1986), Dixit, Grossman, and Helpman (1986)), where there may be multiple principals competing to influence the behavior of an agent, or even multiple agents.[22] The common agency case falls under the three or

[20] This can be substituted for by a carefully constructed continuous function and still give exactly the same incentives.

[21] Of course, this first-best conclusion is due to the absence of any asymmetry of information and hence moral hazard.

[22] Prat and Rustichini (1999) study situations with several principals and several agents, where principals may make strategy contingent payments to the agents who then play a game; and where agents only care about the transfers. Principals may always be seen as players with degenerate strategy spaces in the second-period game. That analysis does

more player case that is analyzed in the next section.

Cournot Duopoly

Consider a classic Cournot duopoly where the action $x_i \in I\!R_+$ is quantity choice of firm i, inverse demand is linear $(a - \sum_i x_i)$ and costs are zero. The payoff function to firm i is $v_i(x_1, x_2) = (a - \sum x_i)x_i$. In this case, the symmetric Cournot equilibrium quantities are $x_i^n = a/3$, and the resulting payoffs are $v_i(x^n) = a^2/9$. If the firms were to collude efficiently,[23] they would choose the monopoly output, $x_1 + x_2 = a/2$ and share the monopoly profits. Thus $\sum \overline{u}_i = a^2/4$.

Let us check that no such pair of strategies and split of monopoly profits is supportable. Consider player 1's solo payoff, u_1^s. If Player 1 chooses output x_1' she can ensure that 2 does not enter the market by the paying contingent on x_1, the minimal transfer as defined above, $t_{12}(x_1', 0) = (a - x_1)^2/4$. In addition 1 can credibly commit to play any such action x_1' in a subgame perfect equilibrium by offering to pay $t_{12}(x_1) = a^2$ if $x_1 \neq x_1'$. To calculate a lower bound on player 1's solo payoff u_1^s we solve:

$$\max\ V = (a - x_1)x_1 - (a - x_1)^2/4.$$

The solution is $x_1 = \frac{3a}{5}$ and so $u_1^s \geq \frac{a^2}{5}$. However the symmetric argument applies to player 2 and so $\sum u_i^s = \frac{2a^2}{5} > a^2/4$. Thus, by Theorem 2 the efficient outcome is not supportable. Notice that this is in contrast to the case of Bertrand competition.

Prisoners Dilemma

We remark that the supportability of the efficient outcome depends on the specifics of the payoffs, and that very simple variations in a game can change its properties. For instance, for the version of the public goods contribution game (prisoners' dilemma) below, the efficient actions were not supportable as we saw in Section 2.

	C	N
C	$2, 2$	$-1, 4$
N	$4, -1$	$0, 0$

not quite fit into our setting, as the agents are not allowed to make payments to each other and there are other restrictions on side payments. We discuss this below.

[23]Efficiency in this context is somewhat perverse, as we are focusing only on the incentives of the firms and ignoring the welfare of the consumers. So, the lack of supporting collusion is actually good from society's perspective.

However, the following modification, which has the same strategic properties (it is strictly dominant for each player to play N, while both playing C is efficient) has different supportability properties. Here, it is straightforward to see that the (modified) solo payoffs are 3 for each player, and so C,C is supportable.

	C	N
C	$3,3$	$0,4$
N	$4,0$	$1,1$

Notice that this points out that supportability depends on the cardinal structure of the game, and not simply the strategic structure.

Commons Problems

Commons problems have results that are similar to those of public goods, with inefficiency being pervasive. There the solo payoffs come from paying the other player(s) to have low usage levels of the common resource, but still to have a high level of usage one's self. Generally, the sum of the solo levels exceeds the efficient point and efficient outcomes are not supportable.

6 Three or More Players

In the case of three or more players, it is relatively easier to support outcomes in the two stage process. That is captured in the following theorems, the first of which addresses the issue of survivability.

THEOREM 4 *If $n \geq 3$, then every Nash equilibrium of the underlying game survives.*

The reason for the much more positive outcomes with three or more players, and for instance the contrast between Theorems 1 and 4, is with more than two players it is possible for players to effectively commit themselves not to play certain strategies through the use of transfer functions. For example, consider a player 1 who would like to be able to commit not to playing an action x_1. Player 1 could simply say that he or she will pay some large amount, say M (which is higher than the maximum payoff to any player in the matrix) to each other player if player 1 were to play x_1. In a two person game, player 2 can undo this by simply committing to pay M back to player 1 if player 1 plays x_1. However, in a three person game, player 2 would have to pay $2M$ back to player 1, and is only getting M from player 1, and so now it is prohibitively costly for player 2 to try to undo player 1's commitment. This type of commitment possibility makes supporting desired strategy-payoff combinations much easier. The importance of commitment to strategies dates (at least) to Schelling (1960). In our analysis of three or more player games, any player can essentially become one who holds a bond (via promised transfers contingent on undesired actions being played) thus committing some other players to play certain strategies.

Note that if one wishes to introduce a third party to hold a bond in a two player game, this can be modeled simply by modeling the third party as a third player in the game who has no actions in the game and no payoff other than transfers received (or made). We discuss this in more detail below.

Let us now turn to the question of supportability. We first provide a full characterization. We also offer a very simple set of sufficient conditions.

Let

$$u_i^{ms}(\bar{t}) = \sup_{t_i} \left[\min_{\mu \in NE(\bar{t}_{-i}, t_i)} EU_i(\mu, \bar{t}_{-i}, t_i) \right].$$

Say that $\bar{t}$ supports $(x, \bar{u})$ if

- $x \in NE(\bar{t})$ and
- $U_i(x, \bar{t}) = \bar{u}_i$ for all i.

Theorem 5 $(x, \bar{u})$ *is supportable if and only if there exists a supporting* $\bar{t}$ *such that* $\bar{u}_i \geq u_i^{ms}(\bar{t})$ *for each* i.

The proof of Theorem 5 is a straightforward variation on the proof of Theorem 3. See the proof of Theorem 7 in the appendix for details.

The necessary and sufficient condition in Theorem 5 is more difficult to check than the corresponding condition in Theorem 3, as Theorem 3 shows that with $n = 2$ one only needs to check the condition with respect to the uniquely defined minimal supporting $\bar{t}$. That is no longer the case with more than two players. Nevertheless, the conditions for supportability are a well-defined linear program, and so the condition is still tractable. In any case, we next provide a much more transparent sufficient condition for supportability.

The following theorem gives a sufficient condition for supportability, which illustrates how permissive support is with three or more players.

THEOREM **6** *If $n \geq 3$ and x is a strategy profile and there exists a Nash equilibrium $\widehat{x}$ such that $v_i(x) \geq v_i(\widehat{x})$ for all i, then $(x, v(x))$ is supportable.*

Theorem 6 states that strategy profile that offers a Pareto improvement over some Nash equilibrium payoffs, is supportable. The proof appears in the appendix. The rough intuition is that it is possible to use the Nash equilibrium as a threat point to which players revert if some player does not make the correct supporting offer of transfers in the first stage.

The following example illustrates the power of Theorem 6. It also of interest since it shows how seemingly small restrictions in the set of admissible transfer functions can be critical. In particular we show how the analysis of common agency of Prat and Rustichini (1999) contrasts with what Theorem 5 predicts for a common agency example, and how this hinges on the set of admissible transfer functions.

EXAMPLE **5** *Efficiency in a Common Agency Example.*

Consider a setting with two principals and two agents. The agents are the only players who take actions. Let us label these as players 1 and 2. The principals are the only ones whose payoffs depend on the play of the game.

	L	R
U	$0, 0, 3, 0$	$0, 0, 0, 2$
D	$0, 0, 0, 2$	$0, 0, 2, 0$

So in this setting player 1 (an agent) takes an action up or down, while player 2 (also an agent) takes an action either left or right. The agents' payoffs are always 0. Player 3 is a principal and would rather that the agents play (U,L) or (D,R), and player 4 is a principal who would rather that the agents would play (U,R) or (D,L).

Theorem 6 shows that the efficient strategy pair (U,L) can be supported together with payoffs (0,0,3,0), since this is in fact an equilibrium of the game with no transfers.[24]

For the above example, Prat and Rustichini's (1999) results conclude that efficiency is not an equilibrium outcome and that the principals would

[24]In fact, Theorem 4 could also be applied and note that all equilibria survive here.

play mixed strategies in the contracts they offer as Prat and Rustichini show in their matching pennies example. The key difference is that Prat and Rustichini only consider contracts between principals and agents, but not between different principals or between different agents. In the contracts that support efficiency in this example and underlie the proof of Theorem 6, there are transfers made off the equilibrium path between agents and/or principals, as a variety of transfer functions work.

As a simple, but useful corollary of Theorem 6, note that in any symmetric game that has a pure strategy Nash equilibrium, a symmetric efficient strategy profile will be supportable.

COROLLARY **1** *If $n \geq 3$ and the game has a pure strategy Nash equilibrium with symmetric payoffs, then any efficient strategy that results in symmetric payoffs is supportable.*

This corollary applies to the symmetric public goods game, commons games, and Cournot games for which such efficient outcomes were not supportable when $n = 2$.

7 Comparisons of Multiple versus Two Players

As is clear from our results, there are differences between the consequences of the results for three or more players and those for two players. Let us offer two important observations in this regard.

Dummy Players and Bonding

We now show how Theorem 6 is also useful in helping to understand how two players might use a third player as a bonding agent.

Consider a two person game (X_1, X_2, v_1, v_2). Let us say that we add a *dummy player* if we add a player with a degenerate singleton action space $X_3 = \{x_3\}$ and with $v_3(x) = 0$ for all $x \in X$.

COROLLARY **2** *If in a two person game there exists an efficient action pair (x_1, x_2) and a Nash equilibrium $(\widehat{x}_1, \widehat{x}_2)$ such that $v_i(x) \geq v_i(\widehat{x})$ for $i \in \{1, 2\}$, then if a dummy player is added to the game, x_1, x_2, x_3 is supportable together with $(\overline{u}_1, \overline{u}_2, 0) = (v_1(x), v_2(x), v_3(x))$ in the three person game.*

Note that the use of the third player in the corollary could also be viewed as placing deposits in escrow to be conditionally returned depending on the actions taken.[25]

Coalitional Considerations

We have seen results for three or more players show that the strategic aspects of side contracts are critically dependent on the number of players, and in particular, differ dramatically depending on whether there are two or more than two players. Part of the reason for the difference is that we have not considered coalitional deviations. If instead of Nash and subgame perfect equilibrium, we considered strong equilibrium and strong perfect equilibrium, then the reasoning behind the three or more player case would look more like the two player case. That is, collectively any coalition of $n-1$ players could always undo the transfers of any other player and then maximize its payoffs subject to only controlling the remaining player through promised transfers. This would result in benchmarks that are similar to the solo payoffs for each coalition of $n-1$ players. In many contexts, this would again lead to combinations of coalitional payoffs that exceed the total efficient payoff in the game.

8 Discussion

We have characterized the outcomes of games that are supportable when players can commit to making strategy contingent side payments to other players. Some basic conclusions from the results can be summarized as follows.

- The incentives to use side payments to affect the strategic aspects of the game are subtle, and at times conflict with efficiency.
- In some cases, efficient strategies that are equilibria in a game without side payments do not survive when side payments are introduced.

[25] Such a device is discussed by Dutta and Radner (2001) as a means of partly solving a commons problem associated with investing in technological development related to slowing global warming.

- The solo payoffs (where only one player can make transfers) are key benchmarks in understanding what outcomes are supportable in games with side payments.
- With three or more players side payments allow for a sort of commitment to strategies that makes supporting efficient strategies (and others) easier to support than with only two players.

While our results suggest that efficiency may not always be supportable, we wish to emphasize an alternative interpretation of the results. The idea that individuals should be able to make transfers that provide incentives to reach efficient points is compelling on the surface. What we have shown here is that carefully analyzing this simple intuition leads to deeper problems and hurdles. Individuals will try to manipulate behavior to their own advantage. This does not preclude the possibility that efficient outcomes can be supported, but it does show that this possibility will be sensitive to the specifics of how individuals can contract before the game. The sorts of simple unilateral offers for compensations that we have analyzed will not suffice, and it will be necessary to have some more complicated forms of bilateral or multilateral contracting with some commitment that excludes further unilateral contracts (that might undo some aspects of the multilateral contract).

Let us discuss some of the restrictions on the types of side payments we have considered and how robust the results should be to changes.

Timing

In our analysis, we have considered only the simultaneous determination of side payment contracts. Let us argue that this is largely inconsequential. Suppose instead that players can move sequentially, and perhaps more than once. As long as a player can respond to the others' contracts, the (modified) solo payoffs are still relevant. Thus, if we end at any equilibrium, it must be that each player is still receiving at least their modified solo payoffs. This leads to a direct extension of our results.

Thus, in order for timing to really be an issue it must either be that some players are restricted not to be able to respond to the contracts of others; or else there must be some frictions in timing, for instance in the form of time discounting and some time or effort cost to writing contracts. But note that neither of these situations should improve efficiency. In the case where some players are restricted in contracting, the game then takes on a Stackelberg flavor. The specific characterization will change depending on how contracting is restricted, but much in the same way that Stackelberg competition is different from Cournot competition while still leading to inefficiency. In the case where one introduces time or other frictions in contracting, contracting is necessarily costly and so efficiency cannot be achieved since it would then require an absence of contracts and (immediate) play of an efficient action profile in the underlying game.

Bilateral Contracting

The contracts we have considered are offered unilaterally — the agents do not ever come together for bilateral or multilateral bargaining in the side contracting phase. Note however, that some of the intuition we have developed here already has some important implications for such a bilateral bargaining setting. After a bilateral contract is signed, an agent may still have an incentive to make a unilateral offer that effectively undoes important aspects of contract and pushes things in (inefficient) directions that are to his or her advantage. Completely eliminating this problem could be done by allowing agents to come together and write a contract that says "no other contracts involving these parties are possible." Our analysis suggests that such exclusionary contracts would be helpful in reaching efficiency, as otherwise agents might make unilateral promises undoing aspects of bilateral contracts.

Another approach would be to consider signing multilateral contracts that effectively rewrite the payoffs in the game (through some transfers) to one that would then be immune to any further contracting. Our results are helpful in understanding how this might be done. For instance, in the context of two person games, if through some initial bilateral contract the payoffs are rewritten so as to turn the game into a pure coordination game[26] In that case, the solo payoffs become those from the efficient point(s) (which are now equilibria) and so efficiency survives. While this is reassuring, one then has to carefully model the bilateral bargaining phase to try to understand whether this contract or some other contract would emerge. The answer is not so obvious, especially given that agents realize that they may still have unilateral offers at their discretion, which endogenizes the bargaining threat point.

[26]A simple recipe for this is to split the total payoff from every profile of actions evenly, or according to some fixed weights.

In any case, our results may be thought of as showing that it is critical to consider more complicated forms of bargaining and contracting in order to support efficient outcomes. This provides a rich agenda for further analysis.

Negative Transfers

Another important aspect of our analysis is that players can only promise positive payments to other players, and cannot make threats of violence (perhaps at a cost to all players) or steal from or tax other players.[27] Our results that efficiency is not always supportable with reward based contracts suggests that threats might be useful in reaching efficiency in some cases, which might partly explain their use.

Let us make an important observation about the robustness of positive transfers versus negative ones. The type of positive transfer contracts that we have considered here are immune to renegotiation ex post (after the second stage game has been played). After the action has been played, since these contracts only involve transfers from one player to others, there are no other transfers that all players would like to renegotiate to, ex post. Violence (and even stealing) will generally be costly for the player inflicting the negative transfer, and so ex post it may be that all players can benefit from a renegotiation. So, in short, allowing for threats of violence, stealing, punishments, etc. contingent on actions might be a useful additional tool for supporting efficient outcomes, but further study is needed and this will involve some attention to ex post renegotiation that was not needed in the analysis here.

Contracts on Contracts

There are two other aspects of the contracting that deserve further attention.

First, the contracts that we have considered are not contingent on the contracts offered by other players. Allowing for such contingencies presents substantial technical hurdles in modeling, as when each contract is contingent upon the form of the other it results in a self-referential problem. This was first pointed out in the competing mechanisms literature (see McAfee (1993), Peck (1995), and Epstein and Peters (1999)). Considering the impact of such contingent contracts is an important open and difficult problem in many contexts. As one can see from Epstein and Peters (1999), it has been

[27] See Schelling (1960) for some interesting discussion of the role of such threats.

a challenge even to prove that problems involving such contingencies are well-posed! A reasonable conjecture (based in part on the understanding of modeling that comes from Epstein and Peters (1999)) is that we might consider contracting on a game with an augmented action space (some $M \times X$, where M is derived endogenously and incorporates some aspects of the contracting but is payoff irrelevant in the second-stage game). In that case, the basic results we have here would still go through, as the solo payoffs would be unchanged. While this seems to be a reasonable conjecture, it appears to be difficult to prove.

The second issue related to contracts on contracts is viewing additional contracting stages before the larger game we have examined here. That is, one might also think of the two stage process that we have considered here as a game, and then consider contracting before it. Of course, one can then build this up arbitrarily.[28] One possibility is to look for some level at which no further contracting would matter. Again, this provides an agenda for future research.

Looking to Mechanism Design and Implementation

We close by noting that our results also have important implications for the mechanism design and implementation literatures. Our results on the survivability of equilibria show that if the mechanism designer cannot control the side contracting of agents, then even if the mechanism is implementing efficient outcomes (when no side contracting is considered), the agents will have incentives to alter the workings of the mechanism through side contracts. Understanding the implementation problem in this broader context could provide very different conditions for implementability. It also raises questions such as which sorts of mechanisms are least susceptible to being undone by side payments. As such side contracting is available (and observed) in many situations, our results here suggest that this is an essential next step in the mechanism design and implementation literatures.

[28]See Lagunoff (1992) for such an approach in the context of selecting mechanisms.

References

Admati, A.R. and M. Perry (1991) "Joint Projects without Commitment," *Review of Economic Studies*, 58, pp. 259–276.

Aghion, P. and P. Bolton. (1987) "Contracts as Barriers to Entry" *American Economic Review*, 77(3), pp. 388-401.

Allen, B. (2001) "Supermechanisms" mimeo: University of Minnesota.

Anderlini, L. and L. Felli (2001) "Costly Bargaining and Renegotiation," *Econometrica*.

Andreoni, J. and H. Varian (1999) "Preplay Contracting in the Prisoners Dilemma," *Proceedings of the National Academy of Sciences*, 96, pp 10933-10938.

Bag, P.K. and E. Winter (1998) "Simple Subscription Mechanisms for Excludable Public Goods," forthcoming in *Journal of Economic Theory*.

Bagnoli, M. and B.L. Lipman (1987), "Provision of Public Goods: Fully Implementing the Core through Private Contributions," *Review of Economic Studies*, 56, pp. 582–602.

Baliga, S. and T. Sjöström (1995), "Interactive Implementation" *Games and Economic Behavior*,27,pp. 38—63

Barbera, S. and M.O. Jackson (2000), "Choosing how to Choose: Self-Stable Voting Rules," mimeo: Caltech.

Bensaid, B. and R.J. Gary-Bobo (1996), "An Exact Formula for the Lion's Share: A Model of Pre-Play Negotiation," *Games and Economic Behavior*, 14, pp 44-89.

Bernheim, B.D. and M.D. Whinston (1986) "Menu Auctions, Resource Allocation, and Economic Influence," *Quarterly Journal of Economics*, Vol. 101, pp. 1-31.

Bucholz, W., K.A. Konrad, and K.E. Lommerud (1997), "Stackelberg Leadership and Transfers in Private Provision of Public Goods," *Review of Economic Design*, 3, pp. 29–44.

Coase, R.H. (1960), "The Problem of Social Cost," *The Journal of Law and Economics*, 3, pp. 1–44.

Danziger and Schnytzer (1991) "Implementing the Lindahl Voluntary-Exchange System," *European Journal of Political Economy*, 7, pp 55-64.

Dixit, A., G. Grossman, and E. Helpman (1986) "Common Agency and Coordination: General Theory and Application to Governmental Policy Making," *Journal of Political Economy*, Vol. 105, pp. 752–769.

Dutta, P. and R. Radner, (2001), "Global Warming and Technological Change," mimeo.

Ellis, C.J. and A. van den Nouweland (1998) "Designing a Mechanism for Inducing Cooperation in Non-Cooperative Environments: Theory and Applications," mimeo: University of Oregon.

Epstein, L. and M. Peters (1999) "A Revelation Principle for Competing Mechanisms," *Journal of Economic Theory*, Vol. 88, pp 119–160.

Fershtman, C., K. Judd and E. Kalai (1991) "Observable Contracts, Strategic Delegation, and Cooperation," *International Economic Review*, Vol. 32, pp 551-559.

Fershtman, C. and S. Nitzan (1991) "Dynamic Voluntary Provision of Public Goods," *European Economic Review*, 35, pp. 1057–1067.

Guttman, J.M. (1978) "Understanding Collective Action: Matching Behavior," *American Economic Review*, 68, pp 251-255.

Guttman, J.M. (1987) "A Non-Cournot Model of Voluntary Collective Action," *Economica*, 54, pp 1-19.

Guttman, J.M. and A. Schnytzer (1992), "A Solution of the Externality Problem Using Strategic Matching," *Social Choice and Welfare*, 9, pp. 73–88.

Hurwicz, L. (1994) "Economic Design, Adjustment Processes, Mechanisms and Institutions," *Economic Design*, 1, pp. 1–14.

Jackson, M.O. (2001) "A Crash Course in Implementation Theory," *Social Choice and Welfare*, 18, pp 655-708.

Jackson, M.O. (2002) "Mechanism Theory," forthcoming in the *Encyclopedia of Life Support Systems*.

Jackson, M.O. and H. Moulin (1992) "Implementing a Public Project and Distributing its Cost," *Journal of Economic Theory*, 57, pp 125–140.

Kalai, E. (1981) "Preplay Negotiations and the Prisoners' Dilemma," *Mathematical Social Sciences*, 1, pp 375–379.

Kalai, E. and D. Samet (1985) "Unanimity Games and Pareto Optimality," *International Journal of Game Theory*, 14, pp 41–50.

Lagunoff, R. (1992) "Fully Endogenous Mechanism Selection on Finite Outcome Sets," *Economic Theory*, 2, pp 462–480.

Marx L. and S.A. Matthews (2000) "Dynamic Voluntary Contribution to a Public Project," *Review of Economic Studies*, 67, pp 327–358.

McAfee, P. (1993), "Mechanism Design by Competing Sellers," *Econometrica*, 61, pp. 1281–1312.

Miller, N.H. and A.I. Pazgal (2001) "The Equivalence of Price and Quantity Competition with Delegation," mimeo: Kennedy School of Government and Olin School of Business.

Peck, J. (1995) "Competing Mechanisms and the Revelation Principle," mimeo: Ohio State University.

Prat, A. and A. Rustichini (1999) "Games Played Through Agents" mimeo: Tilburg University.

Qin, C.-Z. (2001) "Credible Commitments: Ex Ante Deterrence and Ex Post Compensation," mimeo: U.C. Santa Barbara.

Ray, D. and R. Vohra, (1997) "Equilibrium Binding Agreements," *Journal of economic Theory,* 73, pp. 30-78.

Schelling, T. (1960) *The Strategy of Conflict*, Harvard University Press: Cambridge MA.

Segal, I. (1999) "Contracting with Externalities," *Quarterly Journal of Economics*, 114(2) pp. 337-88.

Varian, H.R. (1994a) "Sequential Provision of Public Goods," *Journal of Public Economics*, 53, pp 165–186.

Varian, H.R. (1994b) "A Solution to the Problem of Externalities when Agents are Well-Informed," *American Economic Review*, 84, pp 1278–1293.

Appendix

Proof of Theorem 1: Asking whether x survives is equivalent to asking whether $(x, v(x))$ is supportable (where $v(x)$ is the vector with i-th entry $v_i(x)$). Since x is a Nash equilibrium of the second-stage game, it follows from the definition of $t^{x,\overline{u}}$ that $t^{x,v(x)} = t^0$. This implies that $u^{ms}(x, v(x)) = u^s$, and then Theorem 1 follows from Theorem 3. ▮

Proof of Theorem 2: We show that

$$u_i^{ms}(x, \overline{u}) \geq u_i^s \tag{1}$$

for any i and $x, \overline{u}$. Given (1), the theorem then follows from Theorem 3.

So let us now show (1). Consider any t_i. Let $\widehat{t}_i = t_i + t_{ji}^{x,\overline{u}}$. It follows that

$$t_{ij}(x') - t_{ji}^0(x') = \widehat{t}_{ij}(x') - t_{ji}^{x,\overline{u}}(x')$$

for every x'. This implies that the net transfers across players are identical under (t_{-i}^0, t_i) and $(t_{-i}^{x,\overline{u}}, \widehat{t}_i)$ and so $NE(t_{-i}^0, t_i) = NE(t_{-i}^{x,\overline{u}}, \widehat{t}_i)$. Thus, for each t_i there exists $\widehat{t}_i$ such that

$$\min_{\mu \in NE(t_{-i}^{x,\overline{u}}, \widehat{t}_i)} EU_i(\mu, t_{-i}^{x,\overline{u}}, \widehat{t}_i) = \min_{\mu \in NE(t_{-i}^0, t_i)} EU_i(\mu, t_{-i}^{x,\overline{u}}, t_i).$$

Since this is true for any t_i, it follows that

$$\sup_{t_i} \left[\min_{\mu \in NE(t_{-i}^{x,\overline{u}}, t_i)} EU_i(\mu, t_{-i}^{x,\overline{u}}, t_i) \right] \geq \sup_{t_i} \left[\min_{\mu \in NE(t_{-i}^0, t_i)} EU_i(\mu, t_{-i}^{x,\overline{u}}, t_i) \right],$$

which establishes (1). ▮

Proof of Theorem 3: Let us first show that if $(x, \overline{u})$ is supportable, then $\overline{u}_i \geq u_i^{ms}(x, \overline{u})$ for each i.

Suppose to the contrary that $\overline{u}_i < u_i^{ms}(x, \overline{u})$ for some i and $(x, \overline{u})$ is supportable. It follows that there exists some t_i such that

$$\overline{u}_i < \min_{\mu \in NE(t_{-i}^{x,\overline{u}}, t_i)} EU_i(\mu, t_{-i}^{x,\overline{u}}, t_i). \tag{2}$$

Let $\overline{t}$ be any set of transfers for which $(x, \overline{u})$ is supported. Note that, as argued in the text, it must be that $\overline{t}_j \geq t_j^{x,\overline{u}}$. Let $\widehat{t}_i = t_i + \overline{t}_{ji} - t_{ji}^{x,\overline{u}}$. It follows that

$$t_{ij}(x') - t_{ji}^{x,\overline{u}}(x') = \widehat{t}_{ij}(x') - \overline{t}_{ji}(x')$$

for every x'. This implies that the net transfers across players are identical under $(\overline{t}_{-i}, \widehat{t}_i)$ and $(t^{x,\overline{u}}_{-i}, t_i)$ and so $NE(\overline{t}_{-i}, \widehat{t}_i) = NE(t^{x,\overline{u}}_{-i}, t_i)$. Thus, from (2) it follows that

$$\overline{u}_i < \min_{\mu \in NE(\overline{t}_{-i}, \widehat{t}_i)} EU_i(\mu, t^{x,\overline{u}}_{-i}, \widehat{t}_i).$$

Let i deviate from $\overline{t}$ and announce $\widehat{t}_i$ in the first stage. It follows from the inequality above that the worst possible continuation payoff in the subgame that follows is better than the expected continuation under $\overline{t}$. This contradicts the fact that $\overline{t}$ was played in the first stage of an equilibrium that supports $(x, \overline{u})$.

Next, let us show that if $\overline{u}_i \geq u^{ms}_i(x, \overline{u})$ for each i, then $(x, \overline{u})$ is supportable, and by $t^{x,\overline{u}}$.

Let us specify equilibrium strategies. In the first stage $t^{x,\overline{u}}$ is played and x is played in the second stage. A full specification of the equilibrium strategies includes specification of what happens off the equilibrium path as follows. If in the first stage player i plays $t^{x,\overline{u}}_i$ and player j plays $t_j \neq t^{x,\overline{u}}_j$, then in the second stage that follows the play is $\mu \in NE(t^{x,\overline{u}}_i, t_j)$ that minimizes $EU_j(\mu, t^{x,\overline{u}}_i, t_j)$ over $\mu \in NE(t^{x,\overline{u}}_i, t_j)$. In a subgame following play of t such that $t_i \neq t^{x,\overline{u}}_i$ and $t_j \neq t^{x,\overline{u}}_j$, select any $\mu \in NE(t)$. To see that this forms a subgame perfect equilibrium, note that by the definition of $t^{x,\overline{u}}$ it follows that if $t^{x,\overline{u}}$ is played in the first stage, then it is an equilibrium to play x in the second stage. So we need only show that there is no deviation away from $t^{x,\overline{u}}$ to $t_j \neq t^{x,\overline{u}}_j$ by some j. It follows from the definition of $u^{ms}(x, \overline{u})$ and our specification of off the equilibrium path behavior that if any player j deviates from announcing $t^{x,\overline{u}}_j$ in the first stage then player j's payoff will be no more than $u^{ms}_j(x, \overline{u})$. Since $\overline{u}_j \geq u^{ms}_j(x, \overline{u})$, it follows that this cannot be an improving deviation. ∎

Proof of Theorem 4: Let $M = 1 + max_{i,x',x''}[v_i(x') - v_i(x'')]$. Fix a Nash equilibrium x of the underlying game. Consider the transfer functions

$$t_{ij}(\widetilde{x}) = \begin{cases} 2M & \text{if } \widetilde{x}_i \neq x_i \\ 0 & \text{otherwise.} \end{cases}$$

Under the above transfer functions it is a strictly dominant strategy for each player i to play x_i, and so x is a unique Nash equilibrium in the second-period game. Specify this behavior on the equilibrium path, and off the equilibrium path choose any Nash equilibrium in the second stage. We need only consider that a deviation to some $\widehat{t}_i$ by a player i is not profitable for i. Such a deviation can only be improving if it leads to play of something other than x_{-i} by other players. (If only i changed actions, then i cannot do better given that x was a Nash equilibrium and $t_{ji}(\widehat{x}) = 0$ when $\widehat{x}_j = x_j$.) First, consider the case where a pure strategy Nash equilibrium $\widehat{x}$ is played in the second stage where $\widehat{x}_j \neq x_j$ for some $j \neq i$. Let there be $k \geq 1$ players $j \neq i$ such that $\widehat{x}_j \neq x_i$, and consider some such j. By playing $\widehat{x}$ player j's payoff is

$$v_j(\widehat{x}) - (n-1)2M + 2Mk + \widehat{t}_{ij}(\widehat{x}).$$

If j plays x_j instead, then j's payoff is

$$v_j(\widehat{x}_{-j}, x_j) + 2Mk + \widehat{t}_{ij}(\widehat{x}_{-j}, x_j).$$

For $\widehat{x}_j$ to be a Nash equilibrium conditional on $\widehat{t}$, this implies that

$$\widehat{t}_{ij}(\widehat{x}) - \widehat{t}_{ij}(\widehat{x}_{-j}, x_j) \geq v_j(\widehat{x}_{-j}, x_j) - v_j(\widehat{x}) + (n-1)2M.$$

Given our definition of M and the fact that $n - 1 \geq 2$, it follows that

$$\widehat{t}_{ij}(\widehat{x}) - \widehat{t}_{ij}(\widehat{x}_{-j}, x_j) > 3M.$$

This implies that $\widehat{t}_{ij}(\widehat{x}) > 3M$. This is true for any j with $\widehat{x}_j \neq x_j$. So, player i's utility in the new equilibrium is at most

$$v_i(\widehat{x}) - k3M + k2M.$$

For $k \geq 1$, the definition of M implies that this expression is less than $v_i(x)$. Thus, the deviation cannot be improving. The above reasoning is extends to the case of mixed strategy equilibria in the second stage by similar arguments applied to each realization in the support of the mixed strategy. ∎

Proof of Theorem 6: Consider x and a Nash equilibrium $\widehat{x}$ such that $v_i(x) \geq v_i(\widehat{x})$ for each i.

Set t as follows.

$$t_{ij}(\widetilde{x}) = \begin{cases} 2M & \text{if } \widetilde{x}_{-i} = x_{-i} \text{ and } \widetilde{x}_i \neq x_i \\ 2M & \text{if } \widetilde{x}_{-i} \neq x_{-i} \text{ and } \widetilde{x}_i \neq \widehat{x}_i \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to verify that $x \in NE(t)$, as if i deviates then i pays M to each other player. To support $(x, v(x))$ have the strategies of the players be to play t in the first stage and x in the second stage. Specify off the equilibrium path strategies as follows. Conditional on a single player i deviating from t to some $\widehat{t}_i$ in the first stage, then play $\widehat{x}$ in the second period if $\widehat{x} \in NE(t_{-i}, \widehat{t}_i)$ and otherwise play the worst Nash equilibrium for i out of $NE(t_{-i}, \widehat{t}_i)$. Conditional on more than one player deviating from t in the first stage, play any Nash equilibrium in the resulting subgame.

To complete the proof of the theorem, we need only check that that no player i can benefit by deviating to some $\widehat{t}_i$ in the first period. If $\widehat{x} \in NE(t_{-i}, \widehat{t}_i)$, then the resulting play will be $\widehat{x}$ and so $t_{ji}(\widehat{x}) = 0$ for all $j \neq i$. Thus, the payoff to i will be $v_i(\widehat{x}) - \sum_{j\neq i} \widehat{t}_{ij}(\widehat{x})$. Since this is less than $v_i(\widehat{x})$, it is less than $v_i(x)$ and cannot be a beneficial deviation. Thus, consider the case where $\widehat{x} \notin NE(t_{-i}, \widehat{t}_i)$ but there is some pure strategy $\widetilde{x} \in NE(t_{-i}, \widehat{t}_i)$. If $\widetilde{x} = x$ then it cannot be a beneficial deviation since $v_i(x) \geq v_i(x) - \sum_{j\neq i} \widehat{t}_{ij}(x)$.

We are left with the case where $\widetilde{x} \neq x$ and $\widetilde{x} \neq \widehat{x}$. Let us first show that it must be that $\widetilde{x}_k \neq x_k$ for at least two players k and j, with the possibility that $k = i$. To see this, suppose to the contrary that $\widetilde{x}_k \neq x_k$ for just one k. Given the definition of t_j for each $j \neq i$, it must be that i is paying at least $(2n-3)M$ to each $j \notin \{i, k\}$ for whom $\widetilde{x}_j \neq \widehat{x}_j$ as otherwise j would rather play $\widehat{x}_j$. The transfers to i from each such j amount to at most M and are 0 from any other j. i also gets at most M from k. Thus, by the definition of M, this cannot be a beneficial deviation for i unless $x_{-i,k} = \widehat{x}_{-i,k}$. If $k = i$, then it must be that $x_{-i} = \widehat{x}_{-i}$ and $t_j(\widetilde{x}) = 0$ for all $j \neq i$. Since $\widehat{x}_i$ is a best response to $\widehat{x}_{-i}$ it follows that $v_i(\widehat{x}) \geq v_i(\widetilde{x})$, and so $v_i(\widehat{x}) \geq v_i(\widetilde{x}) - \sum_{j\neq i} \widehat{t}_{ij}(\widetilde{x})$, which implies that this could not be a profitable deviation. Therefore, the only such k must be some $k \neq i$, and thus $\widetilde{x}_{-k} = x_{-k}$. Thus, by the structure t_k for this to be a best response i must pay k at least $(2n-3)M$ and gets M from k and 0 from other j's (for whom $\widetilde{x}_j = \widehat{x}_j$ as shown above). This cannot be profitable for i.

Thus we know that $\widetilde{x}_k \neq x_k$ for at least two distinct players, with the possibility that $k = i$. This means that $\widetilde{x}_{-j} \neq x_{-j}$ for each $j \neq i$ and so by the argument above we know that $\widetilde{x}_j = \widehat{x}_j$ for each $j \neq i$ in order for this to be a profitable deviation for i. This means that $t_j(\widetilde{x}) = 0$ for each $j \neq i$. However, then since $\widehat{x}_i$ is a best response to $\widehat{x}_{-i}$ it follows that $v_i(\widehat{x}) \geq v_i(\widetilde{x})$, and so $v_i(\widehat{x}) \geq v_i(\widetilde{x}) - \sum_{j\neq i} \widehat{t}_{ij}(\widetilde{x})$, which implies that this could not be a profitable deviation.

The extension to the case where in place of $\widetilde{x}$ there is a mixed strategy equilibrium is a straightforward extension of the above reasoning, working on the payments that are made in each realization of the support of the Nash equilibrium. ∎

Games with Continuum Actions

The major technical hurdle faced when the second-period game has infinite (pure) strategy spaces is finding the existence of a subgame perfect equilibrium in the two stage game. If discontinuous transfer functions are allowed (even off the equilibrium path!) then there will be some subsequent subgames where no equilibrium exists. This presents a difficulty, as even restricting attention to continuous transfer functions is then a problem, as it will not be a closed space. One must limit attention to some compact and convex set of transfer functions, for which there always exist second stage equilibria. With this approach, we describe here how the characterization theorems presented above hold in the continuum case.

Consider a game where X_i is a compact metric space and let $\Delta_i(X_i)$ denote the Borel measures on X_i. Let v_i be continuous on X for each i. Consider the set of continuous transfer functions $T = \times_i T_i$.[29] Thus, $NE(t)$ is nonempty and compact for each t.[30]

As in the finite case, define

$$u_i^{ms}(\bar{t}) = \sup_{t_i \in T_i} \left[\min_{\mu \in NE(\bar{t}_{-i}, t_i)} EU_i(\mu, \bar{t}_{-i}, t_i) \right].$$

Note that $\min_{\mu \in NE(\bar{t}_{-i}, t_i)} EU_i(\mu, \bar{t}_{-i}, t_i)$ is well-defined since $EU_i(\mu, \bar{t}_{-i}, t_i)$ is continuous and linear in μ, and $NE(\bar{t}_{-i}, t_i)$ is nonempty and compact.

Say that $\bar{t} \in T$ supports $(x, \overline{u})$ if

- $x \in NE(\bar{t})$ and
- $U_i(x, \bar{t}) = \overline{u}_i$ for all i.

We find the following theorem that covers any n.

[29] One could use a more general space. Any space T for which $NE(t)$ is nonempty and closed (it is then necessarily compact given the space of mixed strategies) will work. In that case one may need to replace $\min_{\mu \in NE(t)} EU_i(\mu, t)$ in the definition of u_i^{ms} with inf, and make some corresponding adjustments in the proof of Theorem 7.

[30] In that case, $U_i(x, t)$ is continuous in x for each i, and then $EU_i(\mu, t)$ is continuous and quasi-concave (in fact linear) in μ. Then by a theorem of Debreu, Fan, and Glicksberg (e.g., see Fudenberg and Tirole (1991)) there exists a Nash equilibrium of the game with t fixed. Closure of the set of Nash equilibria (using weak convergence) then follows easily from the continuity of $U_i(x, t)$ in x.

THEOREM **7** $(x,\overline{u})$ *is supportable if and only if there exists a supporting* $\overline{t} \in T$ *such that* $\overline{u}_i \geq u_i^{ms}(\overline{t})$ *for each* i.

Proof of Theorem 7: Let us first show that if $(x,\overline{u})$ is part of a subgame perfect equilibrium with supporting $\overline{t}$, then $\overline{u}_i \geq u_i^{ms}(\overline{t})$ for each i.

Suppose to the contrary that $\overline{u}_i < u_i^{ms}(\overline{t})$ for some i. It follows that there exists some t_i such that

$$\overline{u}_i < \min_{\mu \in NE(\overline{t}_{-i},t_i)} EU_i(\mu,\overline{t}_{-i},t_i). \tag{3}$$

If player i deviates to play t_i, then for any μ that follows in the subgame, i will benefit. This contradicts the fact that $(x,\overline{u})$ is supported by $\overline{t}$.

Next, let us show that if $\overline{u}_i \geq u_i^{ms}(\overline{t})$ for each i, then $(x,\overline{u})$ is supportable.

Let us specify equilibrium strategies. In the first stage $\overline{t}$ is played and x is played in the second stage. If in the first stage some player i plays $t_i \neq \overline{t}_i$, then in the subgame that follows the play is $\mu \in NE(\overline{t}_{-i},t_i)$ that minimizes $EU_i(\mu,\overline{t}_{-i},t_i)$. In any other subgame select any μ. To see that this forms a subgame perfect equilibrium, note that by the support of $(x,\overline{u})$ by $\overline{t}$ it follows that if $\overline{t}$ is played in the first stage, then it is an equilibrium to play x in the second stage. So we need only show that there is no deviation away from $\overline{t}$ to $t_i \neq \overline{t}_i$ by some i. It follows from the definition of $u^{ms}(\overline{t})$ and our specification of off the equilibrium path behavior that if any player i deviates from announcing $\overline{t}_i$ in the first stage, then player i's payoff will be no more than $u_i^{ms}(\overline{t})$. Since $\overline{u}_i \geq u_j^{ms}(\overline{t})$, it follows that this cannot be an improving deviation. ∎